Computing loop corrections by message passing

A. Ramezanpour

Physics Department and Center for Computational Sciences,

Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy

(Dated: November 30, 2012)

Abstract

Any spanning tree in a loopy interaction graph can be used for communicating the effect of the loopy interactions by introducing messages that are passed along the edges in the spanning tree. This defines an exact mapping of the problem on the loopy interaction graph onto an extended problem on a tree interaction graph, where the thermodynamic quantities can be computed by a message-passing algorithm based on the Bethe equations. We propose an approximation loop correction algorithm for the Ising model relying on the above representation of the problem. The algorithm deals at the same time with the short and long loops, and can be used to obtain upper and lower bounds for the free energy.

I. INTRODUCTION

Interesting problems are usually computationally hard and it is always useful to have efficient and accurate approximation algorithms. Computing local probability marginals for an arbitrary Gibbs measure is one of these fundamental problems in statistical physics and computer science. For example, having an accurate estimation of the probability marginals is enough to solve a constrained satisfaction problem by a decimation algorithm.

There are, of course, some exactly solvable models that can be served as starting points for studying the nontrivial corrections in more interesting and complicated problems. Two well known examples are the mean-field (MF) solution of an infinite dimensional system [1], and the Bethe solution of interacting systems that possess a tree structure [2]. In the former case, we are concerned with the finite dimensional corrections and in the latter, which is the subject of this study, we are concerned with the loop corrections.

Given a tree interaction graph, the local probability marginals can be computed by the belief propagation (BP) algorithm, which minimizes the Bethe free energy by passing messages (or cavity marginals) along the graph edges [3–5]. The loopy belief propagation algorithm is an extension of the BP algorithm to loopy interaction graphs to find a local minimum of the Bethe free energy, which could be larger or smaller than the exact free energy; see Refs. [6–8] for some attempts to construct a convex free energy approximation and ensure the algorithm convergence. Nevertheless, the main strategy to deal with the loops is to group the variables in larger regions to eliminate some loops of given length scale as in the cluster variational method [9, 10] or generalized BP [11]. In the extreme limit we have the junction tree method [12], where the regions are chosen large enough to get a tree interaction graph. There are other algorithms that try to approximate a probability measure with simpler ones having tree structures [13–15].

The local marginals provided by the loopy BP algorithm are not necessarily consistent beyond the two-point correlations, and one way to improve the algorithm is to increase the range of consistent marginals [16, 17]. On the other hand, any fixed point of the loopy BP equations can be used to obtain a loop expansion of finite but exponential number of terms, starting from the loopy BP contribution [18]. In particular, this allows to show that for a class of attractive (or ferromagnetic) models the loopy BP algorithm provides an upper bound for the free energy [19]. Finally, there are some efforts to construct field theories

expanded around the Bethe solution [20, 21].

Here we present an approximation loop correction algorithm that is based on the following observation: In loopy graphs, a global quantity can be computed locally by decomposing the computation into smaller ones distributed among different elements in the graph and collected by messages that are passed along the edges of any spanning tree [22, 23]. Consider the Ising model on an arbitrary graph and one of its spanning trees. Given a spin configuration, one can provide to each spin the effective field originated from the loopy interactions by passing some messages through the spanning tree. The messages are updated at each node to collect the effective fields coming from different parts of the tree; see Fig. 1. In this way, we obtain an exact mapping of the spin configurations on the loopy graph onto the larger space of the spin and effective field configurations on the spanning tree. Of course, this mapping does not change the problem complexity but it offers some approximation loop correction algorithms relying on the above representation.

In the following we discuss more about the details and write the loop correction equations for the Ising model. Then, we study some approximations to reduce the algorithm complexity by considering only a relevant subset of the loopy interactions and treating the other ones in a mean-field approximation. This provides an upper bound for the free energy. We also obtain a lower bound for the free energy by a convex combination of the loopy interactions.

II. LOOP CORRECTIONS IN THE ISING MODEL

Consider the Ising model with Hamiltonian $H = -\sum_{i=1}^{N} B_i s_i - \sum_{(ij)\in E} J_{ij} s_i s_j$ on the interaction graph $G = (V, E)$ with the set of nodes V and edges E. Given a spanning tree $\mathsf{T} = (V, E_0)$ and spin configuration $\underline{s} \in \{-1, +1\}^N$, we rewrite the Hamiltonian

$$H = -\sum_{i=1}^{N} (B_i + B_i^{\mathsf{T}}) s_i - \sum_{(ij)\in E_0} J_{ij} s_i s_j, \tag{1}$$

with $B_i^{\mathsf{T}} \equiv \frac{1}{2} \sum_{j\in\partial i\setminus\partial_0 i} J_{ij} s_j$. Here ∂i and $\partial_0 i$ denote the neighborhood set of i in G and T, respectively. The set $E \setminus E_0$ defines the set of loopy interactions with respect to the spanning tree T. We are going to write the local fields B_i^{T} in terms of some messages that are propagated through the spanning tree. For each loopy interaction $(lk) \in E \setminus E_0$ we introduce messages $h^k_{i\to j}, h^l_{j\to i}$ on edges (ij) in the unique path $l \leftrightarrow k = (l, \ldots, i, j, \ldots, k)$ connecting l and k on T; see Fig. 1. In general, the messages on edge $(ij) \in E_0$ satisfy the

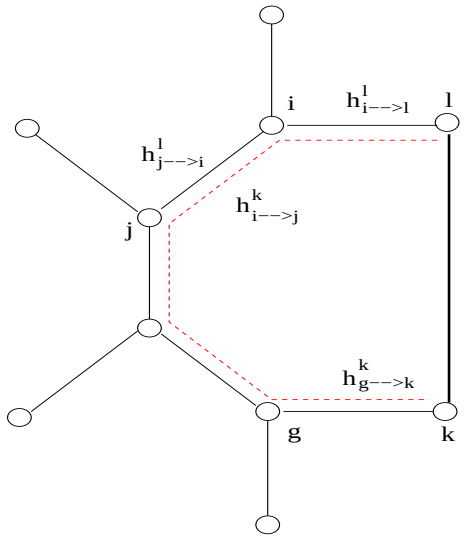

FIG. 1. Propagating the effect of loopy interactions through a spanning tree. Here, the loopy interaction $J_{kl}\sigma_k\sigma_l$ is replaced with the two effective fields $h^l_{i\to l} = 1/2 J_{kl}\sigma_k$ and $h^k_{g\to k} = 1/2 J_{kl}\sigma_l$, which are locally determined by the messages that are passed along the edges in the spanning tree.

following equation:

$$h^l_{i\to j} = \sum_{k\in\partial_0 i\setminus j} h^l_{k\to i} + \frac{1}{2}J_{il}s_i(1-\delta_{l,j}) \equiv \hat{h}^l_{i\to j}. \tag{2}$$

Consequently, we can write $B_i^{\mathsf{T}} = \sum_{j\in\partial_0 i} h^i_{j\to i}$. Note that on each directed edge $(i\to j)\in\mathsf{T}$ we have a vector of messages $\vec{h}_{i\to j} = \{h^l_{i\to j} | l \in \mathsf{T}_{j\to i}\}$. The cavity tree $\mathsf{T}_{i\to j}$ is defined recursively by $i \cup \{\mathsf{T}_{k\to i} | k \in \partial_0 i \setminus j\}$.

Finally, we write the partition function as

$$Z = \sum_{\underline{s}} e^{-H} = \sum_{\underline{s}} \int \prod_{i=1,\ldots,N} \mathbb{I}_h^{(i)} \prod_{j\in\partial_0 i} d\vec{h}_{i\to j} e^{\sum_{i=1}^N (B_i + \sum_{j\in\partial_0 i} h^i_{j\to i})s_i + \sum_{(ij)\in E_0} J_{ij}s_i s_j}, \tag{3}$$

where the indicator function $\mathbb{I}_h^{(i)} \equiv \prod_{j\in\partial_0 i}\prod_{l\in\mathsf{T}_{j\to i}} \delta(h^l_{i\to j} - \hat{h}^l_{i\to j})$ ensures that the sum over the messages $h^l_{i\to j}$ is one when the messages satisfy the equations $\hat{h}^l_{i\to j}$, otherwise it is zero.

Now, the interaction graph is a tree and we can compute the free energy and the local marginals by the Bethe equations. These are self-consistent equations for the cavity marginals $\mu_{i\to j}(s_i; \vec{h}_{ij})$, i.e., the probability of having spin s_i and messages $\vec{h}_{ij} \equiv (\vec{h}_{i\to j}, \vec{h}_{j\to i})$ in absence of node j. The equations governing the cavity marginals, called belief propagation

equations [3], can be written as

$$\mu_{i\to j}(s_i;\vec{h}_{ij}) \propto \int \prod_{k\in\partial_0 i\setminus j} d\vec{h}_{ik} \times \mathbb{I}_h^{(i)} e^{(B_i+\sum_{k\in\partial_0 i} h^i_{k\to i})s_i} \prod_{k\in\partial_0 i\setminus j} \left(\sum_{s_k} e^{J_{ik}s_i s_k} \mu_{k\to i}(s_k;\vec{h}_{ik}) \right). \quad (4)$$

There is, of course, one and only one solution to the BP equations that can be found by iteration starting from the leaves. The Bethe free energy is computed as $F = \sum_i \Delta F_i - \sum_{(ij)\in E_0} \Delta F_{ij}$, where ΔF_i and ΔF_{ij} are the local free energy shifts by adding node i and link (ij) to the interaction graph [5], i.e.,

$$e^{-\Delta F_i} = \sum_{s_i} \int \prod_{k\in\partial_0 i} d\vec{h}_{ik} \times \mathbb{I}_h^{(i)} e^{(B_i+\sum_{k\in\partial_0 i} h^i_{k\to i})s_i} \prod_{k\in\partial_0 i} \left(\sum_{s_k} e^{J_{ik}s_i s_k} \mu_{k\to i}(s_k;\vec{h}_{ik}) \right), \quad (5)$$

$$e^{-\Delta F_{ij}} = \sum_{s_i,s_j} \int d\vec{h}_{ij} e^{J_{ij}s_i s_j} \mu_{i\to j}(s_i;\vec{h}_{ij}) \mu_{j\to i}(s_j;\vec{h}_{ij}). \quad (6)$$

Similarly, one can compute the local marginals $\mu_i(\sigma_i)$ and $\mu_{ij}(\sigma_i,\sigma_j)$.

III. CONSIDERING A SUBSET OF THE LOOPY INTERACTIONS

The time and memory complexity of the above algorithm (in the worst case) grow exponentially with the number of loopy interactions and we have to resort to some reasonable approximations. Here, we focus on a class of approximations that work with a subset of the loopy interactions and preserve the upper bound property of the free energy. The approximation performance then depends on the structure of the spanning tree and the subset of the loopy interactions.

In practice, one can start with a maximum weight $W \equiv \sum_{(ij)\in\mathsf{T}} |J_{ij}|$ spanning tree and add the loopy interactions one by one according to some criterion. Then at each step one obtains an upper bound for the free energy of the original interacting system after adding the energy contribution of the discarded loopy interactions; see Fig. 2. The problem complexity would depend on the number of nonzero vector elements in the $\vec{h}_{i\to j}$. Let us define $X_{i\to j}$ as the set of spins in the cavity tree $\mathsf{T}_{i\to j}$ that interact by loopy interactions with some spins in the cavity tree $\mathsf{T}_{j\to i}$. Then, the vector $\vec{h}_{i\to j}$ takes $2^{|X_{i\to j}|}$ values corresponding to the effective fields for different spin configurations in $\mathsf{T}_{i\to j}$ that are relevant for the spins in $\mathsf{T}_{j\to i}$.

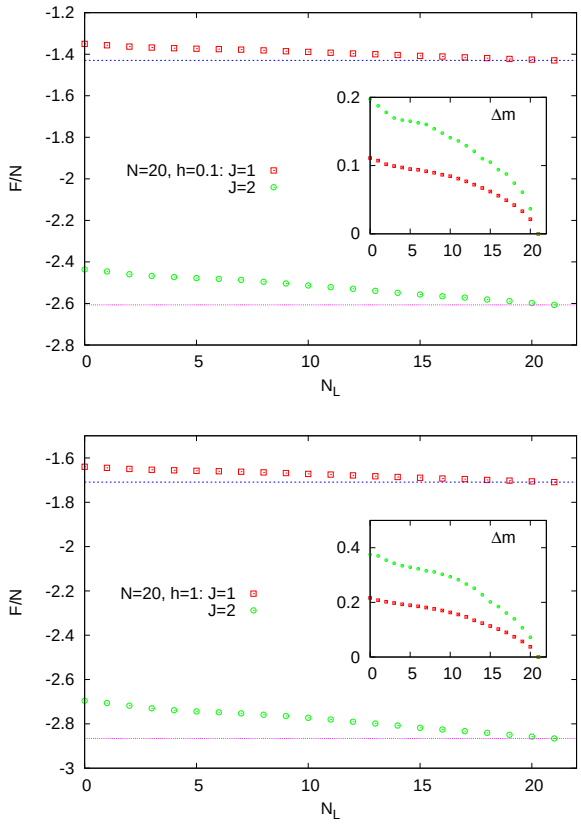

FIG. 2. The exact free energy F and magnetization difference $\Delta m \equiv (\sum_i |m_i^L - m_i^{exact}|)/N$ vs the number of the loopy interactions added to the maximum spanning tree in a $2D$ square lattice of size $N = 4 \times 5$. We take random Gaussian fields B_i of mean zero and variance h^2, and random Gaussian couplings J_{ij} of mean zero and variance J^2. The points are averaged over 500 independent realizations of the random fields and couplings.

Let us start with a mean-field approximation of the loopy interactions, i.e., $H = -\sum_{i=1}^{N}(B_i + \langle B_i^{\mathsf{T}}\rangle)s_i - \sum_{(ij)\in E_0} J_{ij}s_i s_j$, which results to the following BP equations:

$$\mu_{i\to j}(s_i) \propto e^{(B_i + \frac{1}{2}\sum_{k\in\partial i\setminus\partial_0 i} J_{ik}m_k)s_i} \prod_{k\in\partial_0 i\setminus j}\left(\sum_{s_k} e^{J_{ik}s_i s_k}\mu_{k\to i}(s_k)\right). \tag{7}$$

The magnetizations m_i are determined self-consistently by the local marginals

$$m_i = \frac{1}{Z_i}\sum_{s_i} s_i e^{(B_i + \frac{1}{2}\sum_{k\in\partial i\setminus\partial_0 i} J_{ik}m_k)s_i} \prod_{k\in\partial_0 i}\left(\sum_{s_k} e^{J_{ik}s_i s_k}\mu_{k\to i}(s_k)\right), \tag{8}$$

where Z_i is a normalization constant.

Suppose we have partitioned the set of the loopy interactions into two subsets $\mathcal{L}_{BP}$ and $\mathcal{L}_{MF}$ that are to be treated exactly and in the MF approximation, respectively. For example,

the subset $\mathcal{L}_{BP}$ could contain all the loopy interactions within the rth neighborhood of the nodes in T which we denote by $\mathcal{L}_r$. For $r = 0$ we recover the above mean-field approximation. Then, we rewrite the BP equations as

$$\mu_{i\to j}(s_i;\vec{h}_{ij}(r)) \propto \int \prod_{k\in\partial_0 i\setminus j} d\vec{h}_{ik}(r) \times \mathbb{I}_h^{(i)} e^{\left(B_i+\frac{1}{2}\sum_{(ik)\in\mathcal{L}_{MF}(i)} J_{ik}m_k+\sum_{k\in\partial_0 i} h^i_{k\to i}(r)\right)s_i} \times \prod_{k\in\partial_0 i\setminus j}\left(\sum_{s_k} e^{J_{ik}s_i s_k}\mu_{k\to i}(s_k;\vec{h}_{ik}(r))\right), \quad (9)$$

where $\vec{h}_{i\to j}(r)$ is the vector of messages going to the nodes $k \in \mathsf{T}_{j\to i}$ with distance $d_{ik} \le r$. And $\mathcal{L}_{MF}(i)$ is the set of the loopy interactions involving i that are treated in the MF approximation.

Similarly, one can write the self-consistent equations for the magnetization and solve them by iteration starting from random initial cavity marginals and magnetizations. At the fixed point the cavity marginals can be used to find an upper bound for the free energy. There is, of course, no guarantee that the algorithm converges after introducing the MF approximation.

Note that for an arbitrary subset $\mathcal{L}_{BP}$, the computation time scales as $N2^{\chi_{max}}(K_{max}X_{max})^2$ with K_{max} the maximum degree in the spanning tree, X_{max} is the maximum $|X_{i\to j}|$, and $\chi_{max} \equiv \max_i |\sum_{j\in\partial_0 i} X_{j\to i}|$. We can indeed control the algorithm complexity by adding the loopy interactions as long as X_{max} is less than a given value.

In Figs. 3 and 4 we compare the algorithm performance with the loopy BP algorithm in a two-dimensional (2D) square lattice. For the subset $\mathcal{L}_{BP}$ we have chosen the first N_L loopy interactions of largest magnitudes plus the short loopy interactions in the local neighborhood of radius r, such that $|X_{i\to j}| \le X_{max}$ for all the directed links.

IV. A CONVEX COMBINATION OF THE LOOPY INTERACTIONS

One can use the above algorithm to obtain some lower bounds for the free energy. The free energy is a concave function of the fields and the couplings, therefore, $F[\mathbf{B},\mathbf{J}] \ge \sum_g P_g F_g[\mathbf{B}^g,\mathbf{J}^g]$, where P_g is a probability measure and we need to have $\mathbf{B} = \sum_g P_g\mathbf{B}^g$ and $\mathbf{J} = \sum_g P_g\mathbf{J}^g$. The couplings $\mathbf{J}^g$ are chosen such that the free energy F_g can be computed exactly [14, 15]. Let us assume that the interaction graphs G_g defined by $\mathbf{J}^g$ are loopy graphs of complexity less than X_{max} with respect to a spanning tree T_g. Finding the optimal

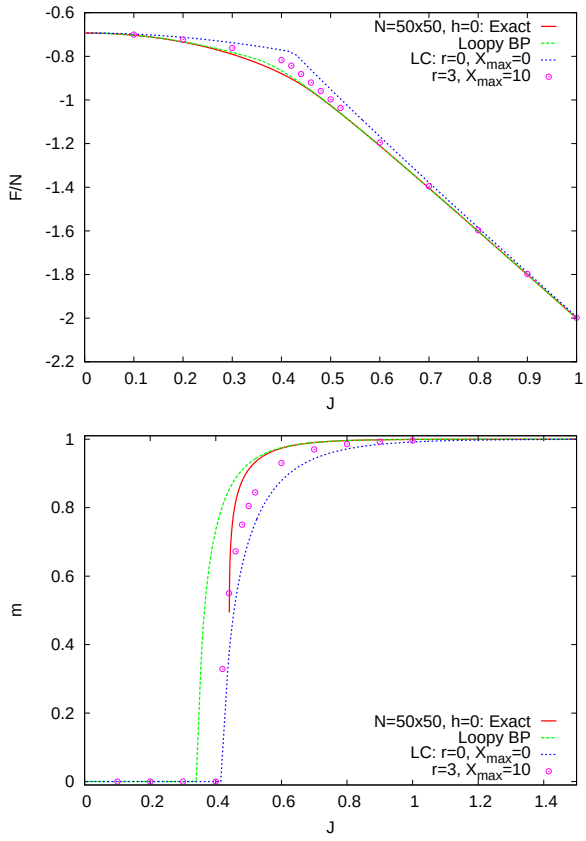

FIG. 3. The free energy F and magnetization m vs the strength of the couplings in a 2D square lattice of size $N = 50 \times 50$ at zero external fields $B_i = 0$ with ferromagnetic couplings $J_{ij} = J$. The loopy BP algorithm and the exact thermodynamic solution are compared with the loop correction (LC) algorithm. The LC results have been obtained with a random spanning tree for different values of r, giving the length scale of the short loops, and X_{max}, giving the computational complexity of the algorithm determined by the structure of the loopy interaction graph.

lower bound is difficult but any consistent set of the couplings and the measure P_g give a lower bound for the free energy. Here, we take the maximum weight spanning tree T, and partition the set of the loopy interactions $\mathcal{L}$ into subsets $\mathcal{L}_r$ and $\{\mathcal{L}_g | g = 1, \ldots, \mathcal{N}\}$, and set $G_g = \mathsf{T} + \mathcal{L}_r + \mathcal{L}_g$. Moreover, we set $\mathbf{B}^g = \mathbf{B}$ and $J^g_{ij} = J_{ij}\delta_{(ij)\in\mathsf{T}+\mathcal{L}_r} + (J_{ij}/P_g)\delta_{(ij)\in\mathcal{L}_g}$, where $P_g = W_g/(\sum_{g'} W_{g'})$ with $W_g \equiv \sum_{(ij)\in\mathcal{L}_g} |J_{ij}|$. Figure 4 displays the lower bounds that we obtain in this way for the free energy in a 2D square lattice with random fields and couplings .

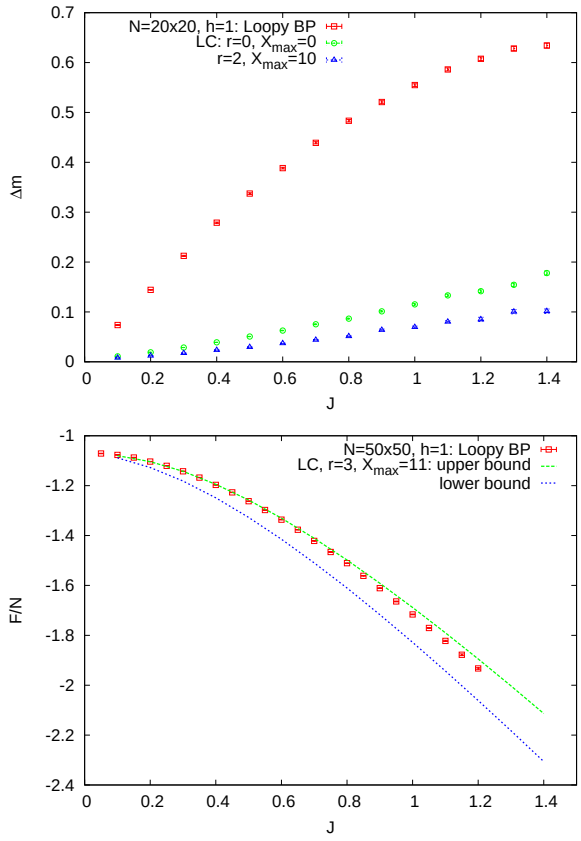

FIG. 4. The magnetization difference Δm (relative to the Monte Carlo results) and the free energy upper and lower bounds vs the strength of the couplings in a 2D square lattice with random Gaussian fields B_i of mean zero and variance h^2, and random Gaussian couplings J_{ij} of mean zero and variance J^2. The loopy BP algorithm is compared with the loop correction (LC) algorithm based on the maximum spanning tree and for different values of r and X_{max}. The data points in the top panel are averaged over at least 10 independent realizations of the random fields and couplings. The lower and upper bounds in the bottom panel have been obtained for a single instance of the problem whereas the loopy BP results are averaged over at least 10 independent realizations.

V. CONCLUSION

In summary, we introduced an approximation message-passing algorithm to compute loop corrections in the Ising model. The approximation works with a relevant subset of the loopy interactions and uses the mean-field approximation to deal with the other loopy interactions. Obviously, the algorithm in this form is more suited to systems with strongly heterogeneous couplings. It would be very useful to have other approximation algorithms which treat all

the loopy interactions in the same manner. And finally, the algorithm can be used to obtain better lower bounds for the free energy than the naive one that we presented here.

ACKNOWLEDGMENTS

I would like to thank A. Bruanstein, M. Mezard, A. Pagnani, and R. Zecchina for helpful discussions. Support from ERC Grant No. OPTINF 267915 is acknowledged.

---

[1] W. L. Bragg and E. J. Williams, Proc. Roy. Soc. A **145**, 699 (1934).

[2] H. Bethe, Proc. Roy. Soc. London A **150**, 552 (1935).

[3] F. R. Kschischang, B. J. Frey, and H. -A. Loeliger, IEEE Trans. Infor. Theory **47**, 498 (2001).

[4] M. Mézard and G. Parisi, Eur. Phys. J. B **20**, 217, 2001.

[5] M. Mézard and A. Montanari, *Information, Physics, and Computation*, Oxford University Press, Oxford, 2009.

[6] A. L. Yuille, Neural computation **14**, 1691 (2002).

[7] T. Heskes, Journal of Artificial Intelligence Research **26**, 153 (2006).

[8] Y. Weiss, C. Yanover, and T. Meltzer, Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence (UAI2007), 2012.

[9] R. Kikuchi, Phys. Rev. **81**, 988 (1951).

[10] A. Plizzola, J. Phys. A: Math. Gen. **38**:R309 (2005).

[11] J. S. Yedidia, W. T. Freeman, and Y. Weiss, IEEE Trans. Infor. Theory **51**(7), 2282 (2005).

[12] S. L. Lauritzen and D. J. Spiegelhalter, Journal of the Royal Statistical Society B **50**(2), 157 (1988).

[13] T. Minka and Y. Qi, Advances in Neural Information Processing Systems 16, Cambridge, MA, 2004. MIT Press.

[14] M. J. Wainwright, T. S. Jaakkola, and A. S. Willsky, IEEE Trans. Infor. Theory, **49**(5), 1120 (2003).

[15] V. Kolmogorov, IEEE Transactions on Pattern Analysis and Machine Intelligence, **28**(10), 1568 (2006).

[16] A. Montanari and T. Rizzo, J. Stat. Mech. (2005) P10011.

[17] J. M. Mooij, B. Wemmenhove, H. J. Kappen, and T. Rizzo, Proc. AISTATS (2007).

[18] M. Chertkov and V. Y. Chernyak, J. Stat. Mech. (2006) P06009.

[19] E.B. Sudderth, M. J. Wainwright, and A. S. Willsky, Proceedings of Neural Information Processing Systems (2007).

[20] K. B. Efetov, Physics A **167**, 119 (1990).

[21] G. Parisi and F. Slanina, J. Stat. Mech. (2006) L02003.

[22] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, 2003.

[23] A. Ramezanpour, J. Realpe-Gomez, and R. Zecchina, Eur. Phys. J. B **81**, 327 (2011).